Imperial College London

SAMGrid Stakeholders Meeting

23rd March

Gavin Davies, Imperial College London

Opening remarks

Imperial College London

- Requests much as presented at previous Stakeholders meeting of 13/12/05 and at GDMs of 10/01/06 & 28/02/06
- DØ reliant on SAMGrid for
  - All data handling, most production computing

→ Strengthen limited data handling support

- Reliance increasing
  - Increased dependence / increased functionality
- Reflected in goals for this year
  - Increased interoperability / automation ⇨ increased production rates along with increased functionality
  - Joint effort between central SAMGrid team and remote sites
- Re-fixing
  - Full 1.4 billion events fixed ahead of schedule – thanks to all
  - Advanced some of longer term goals, esp. interop with LCG & OSG
    - But some consequences

SAMGrid CY2006 Goals

(as defined in Fall 05 and previously presented here / to GDM)

Imperial College London

- Automated production of MC on LCG — Mid Jan
  - System fully tested by contribution to re-fixing
  - Now back to MC production – Collaboration driven
- Demonstration of primary processing with SAMGrid — End Jan
  - Ongoing, some delay from re-fixing
  - Collaboration driven
- Deployment of SAMGrid v7 (inc. d0runjob, MC, fixing etc) — 1st Mar
  - SAMGrid used for re-fixing for 1st time but in v5
  - Needed by both SAMGrid team & expt
  - Big change – effecting many areas - Dedicated project within DØ setup11/05
  - Workplan & timelines developed for key activities
  - Some progress on d0runjob but strong overlap with re-fixing
  - Now major effort from both SAMGrid team & collaboration, as of early Mar
  - Timescale ~ 2 months from now
  - Must deal with this whilst maintaining other activities
  - Will be challenging – in particular maintaining operations
    ⇨ a dedicated "operations/deployment" activity
  - Transition across sites will have to be adiabatic

SAMGrid CY2006 Goals

(as defined in Fall 05 and previously presented here / to GDM)

Imperial College London

- SAM and SAMGrid performance — 1st Jun
  - Includes data handling / operations
  - " Operations/ deployment" team - underway
  - Some 'data-handling' issues seen during refixing (normal DH ops ~0.5FTE)
    - Overall throughput – access to servers, tape, off-site procedures…
    - Not a criticism - "expected" – part of the "operations/deployment" post
- Primary processing with SAMGrid as default — ~~Mid July~~ Mid Jun
  - Will test during shutdown
  - Collaboration driven but will require some SAMGrid effort
  - Looking to include skimming – will require some SAMGrid effort
- Automated production of MC with OSG — End Oct
  - 1st use on OSG with non-native SAMGrid installation carried out for re-fixing
- Strategy for improved user analysis — Mid Nov
  - grid user case?, resilient dcache?
- Automated submission of MC (or reprocessing) by SAM-shifter — Mid Dec

Summary / Conclusions

Imperial College London

- DØ is increasingly reliant on SAMGrid – list of activities reflects this
  - → Support and incremental extensions..
- Full 1.4 billion events fixed ahead of schedule
  - A tremendous success – thanks to all.
  - Advanced some of longer term goals, in particular on interoperability with LCG and OSG.
- An unavoidable consequence was delay to SAM v7 transition
  - Now restarted in earnest
  - Significant effort on both SAM-Grid & collaboration side
    - Timescales similar for both
- SAM v7 transition will leave "operations/deployment" exposed
    - Saw examples of this during re-fixing
    - New dedicated effort on this also covers limited data-handing support
- Progressing on other goals, relying more heavily on collaboration SAMGrid effort during v7 transition.

Imperial College London

Back-up slides

CY2006 Goals

(as defined in Fall 05 and previously presented)

Imperial College London

Use of luminosity database	1st Jan✓
Automated production of MC on LCG	Mid Jan
Demonstration of primary processing with SAM-Grid	End Jan
Sustainable MC capacity of 5M events /week	End Feb
Completion of trigger database taskforce	1st Mar
Deployment of SAM-Grid v7 (inc. d0runjob, MC, fixing etc)	1st Mar
Entire Run IIa data (~1.3fb-1) set processed.	5th Mar
p20 MC certified	19th Mar
p20 MC in production on the farms	9th Apr
CAB operating as part of Fermigrid / OSG	1st May
SAM performance – increased cache metrics/ functionality	1st Jun
SAM-Grid performance – "operations/deployment" (new team*) inc data handing, stability	1st Jun
Primary processing with SAM-Grid as default	Mid July
Automated production of MC with OSG	End Oct
Sustainable MC capacity of 7M events /week	1st Nov
Strategy for improved user analysis - grid user case?, resilient dcache?	Mid Nov
Automated submission of MC (or reprocessing) by SAM-shifter	Mid Dec

* Activity also implicit in increased production rates – to start now.

Deployment Team – Draft Charge

Imperial College London

SAMGrid is being increasingly used by Tevatron experiments, and in particular by the DØ collaboration. DØ is effectively reliant on SAMGrid for much of its production computing. SAM has been used for all data handling since the start of RunII, SAMGrid is now the default for Monte Carlo production and recent developments include the reprocessing and fixing of data. Next steps include the inclusion of other production tasks such as primary processing or skimming and finally individual analysis jobs. In parallel there is a very active programme to make SAMGrid fully interoperable with other grids such as LCG and OSG, to give the Run II experiments access to the resources required to analysis the growing Run II data sets.

In the past the same developers have been responsible for assisting with the installation at remote sites, operational support and optimization of the overall system performance. In view of the increased functionality, increased scale and ongoing development this is no longer tenable. Hence a SAMGrid deployment / operations team has been formed. This group, as a sub-team within the SAMGrid project, will be responsible for:

DRAFT

- Determining and documenting the appropriate deployment procedure for SAMGrid, either as a native installation or as part of the LCG and OSG interoperability projects.
- Assisting with the deployment of SAMGrid at remote sites, either as a native installation or as part of the LCG and OSG interoperability projects.
- Maintenance of the central SAMGrid infrastructure at Fermilab at the appropriate production level.
- Assisting with the operational support of the remote facilities in conjunction with the remote sites themselves.
- Monitoring and optimizing the overall data-handling associated with SAMGrid operations, in conjunction with the experiment's existing data-handling support and the remote sites. This will involve determining whether suitable tools are available and contributing to their development if not.
- The precise nature of these tasks will evolve as the team is established. The team will have a SAM-Grid core, supplemented with a member from each collaboration on a 6 month rotation at Fermilab.